The effect of sodium ion concentration on intrastrand base-pairing in single-stranded DNA

Alan R. Wolfe and Thomas Meehan*

Division of Toxicology and Department of Pharmacy, University of California, San Francisco, CA 94143, USA

Received March 31, 1994; Revised and Accepted June 23, 1994

ABSTRACT

The salt-induced formation of duplex structure (primarily hairpin loops) in denatured calf thymus DNA was monitored by measuring the decrease in absorbance at 260 nm as a function of increasing sodium ion concentration. It was found that this process was noncooperative and could be accurately described by the mass-action expression for the reversible formation of a binary complex: single strand (coil) + free sodium ion $\rightleftharpoons$ hairpin (with associated sodium ion). The equilibrium constant for the transition was found to be $6 \text{ (M Na}^+)^{-1}$. The extrapolated absorbance at infinite salt concentration represents 11% hyperchromicity, which is one third of the hyperchromicity of denatured DNA in the absence of salt (36%).

INTRODUCTION

The secondary structure of a random-sequence single DNA strand is strongly dependent on the counterion concentration. Salt induces the formation of short, imperfect duplexes in single-stranded (ssDNA), which result primarily from short-range, intrastrand base pairing, leading to the formation of hairpins (1–2). The change in secondary structure can significantly affect ligand-ssDNA interactions. Ligand-DNA binding generally becomes weaker as the salt concentration increases, with $-\log K$ increasing linearly with $\log [\text{Na}^+]$. The slope of this plot has been interpreted as representing the number of Na^+ ions thermodynamically bound to DNA that are released per ligand bound (3–4). The ion release occurs as a result of ion pair formation in the case of cationic ligands (3–4), and also as a result of DNA strand extension in the case of intercalators (5).

In experiments with an uncharged intercalating agent, we found that the salt effect on binding affinity was in qualitative accord with the theoretical prediction in the case of native double-stranded DNA (dsDNA), but not in the case of ssDNA, where the binding affinity initially increased with added salt (6). In contrast, the salt effect on interactions between a cationic intercalator and DNA, which are predominately electrostatic, is very similar for dsDNA and ssDNA.) In order to explain the

salt effect on neutral intercalator-ssDNA interactions, it was found necessary to take into account the salt-induced ssDNA conformational transition. Polycyclic aromatic hydrocarbons and their uncharged derivatives bind selectively to duplex regions of ssDNA, so the salt-induced folding of ssDNA effectively increases the number of binding sites. When the calculation of the neutral intercalator-ssDNA binding affinity was corrected for the variation in the number of duplex binding sites, the salt effect was found to be the same as with dsDNA.

MATERIALS AND METHODS

Single-stranded DNA was prepared by boiling a solution of high molecular weight calf thymus DNA (42% G+C content) 30 min in a tightly capped vial. The sample was then quenched in ice water, diluted 1–75 in 2 mM sodium phosphate buffer (pH 7.5) containing 0.1 mM EDTA, and diluted 1–2 in the same buffer containing 0.1 mM EDTA and 0, 1, or 4 M NaCl, yielding denatured DNA solutions (89.6 μM in bases, using $\epsilon_{260} = 6550 \text{ M}^{-1} \text{ cm}^{-1}$) containing 0, 0.5 or 2 M NaCl. The sodium ion concentration of the solution without NaCl was 3.9 mM (the contribution from the DNA was determined by atomic absorption). Cuvettes containing the ssDNA solution without salt were put in the sample and reference positions of a Cary Model 118 spectrophotometer. Aliquots of ssDNA + NaCl solutions were then mixed with the sample solution (to increase the salt concentration while holding DNA concentration constant), and the decrease in absorbance at 260 nm was measured.

RESULTS

The manner in which the 260 nm absorbance of ssDNA varies with NaCl concentration is shown in Figure 1. The extinction coefficient of the denatured DNA without added salt was 35.8% higher than that of native DNA. The drop in the extinction of ssDNA upon addition of salt is immediate, and results from the increased base stacking that accompanies hairpin formation. This conformational change is distinct from the process of renaturation (which is far slower for high molecular weight DNA). Increasing sodium ion concentration favors duplex formation because

*To whom correspondence should be addressed

Figure 1. The effect of sodium ion concentration on the hyperchromicity of ssDNA. The 260 nm extinction of denatured calf thymus DNA in the presence of 2 mM sodium phosphate buffer (pH 7.5), 0.1 mM EDTA and various concentrations of NaCl is given relative to that of native calf thymus DNA (in 10 mM NaCl). The curve was obtained by fitting the data to equation (1) using least squares nonlinear regression ($R = 0.9998$).

counterion shielding of the phosphate charges facilitates both the approach of opposing strands and the contraction that occurs within each strand in the coil-to-helix transition.

The curve in Figure 1 was obtained from the mass-action expression for the formation of a binary complex. The concentration of sodium ions always greatly exceeded that of the DNA phosphates, suggesting that the equilibrium can be analyzed as described in the Appendix (this treatment assumes that cooperativity is absent). Letting s'_a , s'_F , and s'_B represent the apparent (measured) extinction coefficient of ssDNA (ϵ_a) and its extinction coefficients when entirely in the single-stranded coil (ϵ_{ss}) and hairpin states (ϵ_{hp}), respectively, equation (A3) can be rearranged to give:

$$\epsilon_a = \epsilon_{ss} - \frac{(\epsilon_{ss} - \epsilon_{hp})K[Na^+]}{1 + K[Na^+]} \quad (1)$$

([DNA] $\ll$ [Na $^+$])

(in this paper, restrictions on the validity of equations will be noted within parentheses under the equation number). Fitting the data in Figure 1 to equation (1) gives $6.0 \text{ (M Na}^+)^{-1}$ for K , 35.9% hyperchromicity for ϵ_{ss} , and 11.1% hyperchromicity for ϵ_{hp} [% hyperchromicity is $(\epsilon_a/\epsilon_{ds} - 1) \times 100$, where ϵ_{ds} is the extinction coefficient of native DNA]. The values of f_B [= $(\epsilon_a - \epsilon_{ss})/(\epsilon_{hp} - \epsilon_{ss})$], the mole fraction of ssDNA in the hairpin state, ranged from 1% to 84%.

An estimate of the mole fraction of the ssDNA in the duplex state (χ_{ds}) can be obtained from the expression $\chi_{ds} = (\epsilon_a - \epsilon_{ss})/(\epsilon_{ds} - \epsilon_{ss})$. This is based the assumptions that the amount of base stacking in duplex regions of ssDNA is the same as in native DNA, and that salt does not increase the amount of stacking in non-base-paired regions of ssDNA. (Violation of these assumptions would raise or lower the estimate of χ_{ds} , respectively.) When base pairing is prevented with alkali, the latter assumption is at least approximately valid (1). If χ_{ds} is substituted for m in equations (A3)–(A6), L becomes the value of χ_{ds} at infinite salt, i.e. the mole fraction of duplex in the salt-

Figure 2. Duplex formation in ssDNA as a function of sodium ion concentration. The mole fraction of duplex (χ_{ds}) in denatured calf thymus DNA was estimated from the data in Figure 1 as described in the text. The line was obtained by fitting the data to equation (A6) using least squares linear regression ($R = 0.997$). The 0 NaCl (3.9 mM Na^+) data point (not shown) is an outlier in this plot and was not used in the calculation.

induced hairpin state of ssDNA. The linear plot in Figure 2, based on equation (A6), gives the same value for K obtained in Figure 1 (6.1 M^{-1}), and 0.69 for L . The same results were also obtained (by computer) using the nonparametric 'direct linear plot' procedure associated with equations (A7a–b).

DISCUSSION

The excellent fits of the experimental data to equations (A3) and (A6) indicate that sodium ions bind to ssDNA singly and without cooperativity. This means that the formation of multiple hairpins on a single strand is not cooperative, and that cooperativity in the formation of a given hairpin stem generally does not extend beyond the number of base pairs formed in response to binding of a single sodium ion (see below). Since random-sequence DNA was used, the values of K and L obtained must be regarded as averages. However, the linearity of the data plotted in Figure 2 indicates [see equation (A8)] that the distribution of microscopic K_i values is not very broad. (For example, if two types of hairpin-forming sequence with the same $\epsilon_{ss} - \epsilon_{hp}$ values were present in equal abundance, the data would imply that the two K values differed by less than a factor of four.) This allows formulation of an expression based on equation (A3) (substituting χ_{ds} for m) for the duplex content of ssDNA as a function of Na^+ concentration.

The foregoing analysis does not reveal how much duplex is formed per bound sodium ion during hairpin formation in ssDNA. Although studies directly addressing this question have not yet been reported, some insight may be obtained from related work in polyelectrolyte theory. According to the counterion condensation model (3, 7–9), a linear polyanion will attract a high local concentration of condensed counterions when its linear charge density (b^{-1} , where b is the average axial distance per unit of polymer charge) exceeds a critical value (b_{cr}^{-1}). For a polyanion bearing univalent charges in an aqueous solution at room temperature containing univalent cations only, $b_{cr} = 0.714 \text{ nm}$. When the critical charge density is exceeded, the

predicted number of condensed and screening univalent cations thermodynamically bound per unit polymer charge (ψ) is:

$$\psi = 1 - b / (2b_{cr}) \quad (2)$$

$$(b < b_{cr})$$

The same expression can be obtained from the Poisson–Boltzmann cylindrical cell model for the limiting case of very low salt concentration (10, 11). The values of b for duplex B-DNA and single-stranded DNA in the open-coil form are 0.17 and 0.41 nm (9), giving values of 0.88 and 0.71 for ψ_{ds} and ψ_{ss} , respectively. Thus equation (2) predicts that $\Delta\psi$ for the ssDNA→dsDNA transition (the number of sodium ions taken up per phosphate during DNA renaturation, or equivalently, the number released per phosphate during denaturation) is 0.17; this is close to the measured value of 0.15 (12). This indicates that in this transition three base pairs form per sodium ion bound.

End effects limit the extent of counterion binding to oligonucleotides relative to that in polynucleotides. Nevertheless, both experiment and theory indicate that the $\Delta\psi$ values for duplex formation between two single strands are similar for oligo- and polynucleotides, because the end effect on ψ is comparable for single-stranded and double-stranded oligos of the same length. However, the $\Delta\psi$ values (based on all the phosphates present) for hairpin formation in single-stranded oligos are considerably smaller (12). This result arises because folding shortens the oligo and strengthens the end effects. The situation is analogous when hairpins form within a single-stranded polynucleotide: end effects will limit cation binding to the short folded regions (relative to the binding that would occur in long duplex regions), but not to unfolded regions. Thus, the value of $\Delta\psi$ in this process should also be lower than that for single strand dimerization.

Counterion condensation is not a mass-action phenomenon, since ψ is generally insensitive to sodium ion concentration. Over a wide range of $[Na^+]$, there is an equilibrium between the open-coil and hairpin conformations of ssDNA, each of which should be characterized by its own ψ value (ψ_{ss} or ψ_{hp}). Since ψ_{hp} is higher, the principle of Le Chatelier dictates that the hairpin state should be favored at higher $[Na^+]$. It is this shift in the conformational equilibrium that appears to obey the law of mass action.

APPENDIX

A general analysis of heterodimeric physical complex formation when one species is present in large excess

Let it be assumed that two species, α and β , reversibly form a binary complex with association constant K . Their total concentrations will be designated C_T and D , respectively. The fraction of α bound, denoted f_B , is C_B/C_T , where $C_B \equiv$ the concentration of complex or bound α . Assume one can measure a signal s from free α and/or the complex, to which free β does not contribute. If the magnitude of s at constant f_B is dependent on C_T , as would be true for fluorescence intensity or absorbance, the quantity $s' = s/C_T$ will designate the signal normalized to a standard C_T value. (For a non- C_T -dependent signal, such as a binding-induced change in n.m.r. chemical shift under conditions of rapid exchange, $s' = s$.) The signals of free and bound α will be designated s'_F and s'_B . For a given value of D , the apparent value of s' will be denoted s'_a , and the quantity m will be defined as $s'_a - s'_F$. As the concentration of β becomes very large ($D \rightarrow \infty$), α will become fully bound

($f_B \rightarrow 1$), and m will approach its limiting value $L = s'_B - s'_F$. Because this is a two-state system, $s'_a = f_B s'_B + (1 - f_B) s'_F$, giving:

$$f_B = m / L \quad (A1)$$

If s'_B and s'_F are independent of C_T , f_B and D , they will be constants (assuming other solution parameters do not change). This means that L will also be constant, and m will be linearly proportional to f_B . A method to check whether m actually varies linearly with f_B has been described (13).

If β is present in large excess, C_B will be negligible in comparison to D ; the mass-action expression for K can then be written:

$$K = \frac{C_B}{(C_T - C_B)D} = \frac{f_B}{(1 - f_B)D} \quad (A2)$$

$$(C_T \ll D)$$

If α is a polymer, the assumption that K is constant over a range of f_B values implies that the binding of β is noncooperative. Solving equation (A2) for f_B and substituting from equation (A1) gives:

$$m \frac{LKD}{1 + KD} \quad (A3)$$

$$(C_T \ll D)$$

Inverting equation (A3) gives:

$$\frac{1}{m} = \frac{1}{LK} \frac{1}{D} + \frac{1}{L} \quad (A4)$$

$$(C_T \ll D)$$

Multiplying equation (A4) by D gives:

$$\frac{D}{m} = \frac{1}{L} D + \frac{1}{LK} \quad (A5)$$

$$(C_T \ll D)$$

Multiplying equation (A4) by mL and rearranging gives:

$$m = (-1/K) \frac{m}{D} + L \quad (A6)$$

$$(C_T \ll D)$$

Equations (A3)–(A6) are analogs of the Michaelis–Menten, Lineweaver–Burk, Eadie–Hofstee and Hanes–Woolf equations of enzyme kinetics (with $1/K$ corresponding to K_m). Equations (A4)–(A6) are linear transformations of equation (A3), written in the form $y = (\text{slope})x + \text{intercept}$, which allow determination of K and L via linear regression from the appropriate type of plot. Another example of their use is in the analysis of the binding of a dye by DNA (6). The double-reciprocal equation (A4) has tended to be most popular, probably because it results in plots with the highest correlation coefficients. However, when used with unweighted data, the estimates of the unknowns obtained from equations (A5) or (A6) are significantly more accurate (14–15).

If the frequency of outliers exceeds that expected for a normal distribution, a nonparametric method is likely to give estimates of the binding parameters superior to those obtained from a least squares method (16). A simple nonparametric approach (a variant

of the 'direct linear plot') involves plotting a line, based on the points (0, $1/m$) and (D/m , 0), for each data point. Such a line represents all the pairs of ($1/KL$, $1/L$) values consistent with the data point (the relevant equation is obtained by solving equation (A4) for $1/L$). The abscissa and ordinate of the intersection of two lines give the estimates of $1/KL$ and $1/L$, respectively, determined by that pair of data points (17). From n data points one obtains $n(n-1)/2$ pairs of estimates, and the median of all the estimates for a parameter is taken as the best estimate. The point of intersection of lines 1 and 2 is given by:

$$\frac{1}{KL_{(12)}} = \frac{\frac{1}{m_2} - \frac{1}{m_1}}{\frac{1}{D_2} - \frac{1}{D_1}}; \quad \frac{1}{L_{(12)}} = \frac{\frac{D_2}{m_2} - \frac{D_1}{m_1}}{D_2 - D_1} \quad (\text{A7a, b})$$

$$(C_T \ll D)$$

The estimates of K and L are then obtained from the best estimates for $1/KL$ and $1/L$.

Let it now be assumed that the measured signal comes from an assortment of n different types of molecule or binding site, rather than from a single type. The mole fraction of α_j is $\chi_j = C_{Tj} / \sum_{i=1}^n C_{Ti}$, where C_{Ti} denotes the total concentration

of type i . If each α_i is characterized by its own K_i and L_i values, the apparent value of f_B will be:

$$\frac{m'}{L'} = \frac{\sum_{i=1}^n f_{Bi} L_i \chi_i}{\sum_{i=1}^n L_i \chi_i} \quad (\text{A8})$$

where $f_{Bi} = K_i D / (1 + K_i D)$ (if $\sum_{i=1}^n C_{Ti} \ll D$). If the various K_i values differ, the plots of m' and D based on equations (A4)–(A6) will not be linear. The linear relationship between m' and the real value of f_B ($= \sum_{i=1}^n f_{Bi} \chi_i$), however, will be maintained unless there is variation in both the K_i and L_i values.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the National Cancer Institute (CA 40598), the Elsa U. Pardee Foundation, and the UCSF Academic Senate. We thank Andrea Mazel for assistance in preparation of the manuscript.

REFERENCES

- Studier, F.W. (1969) *J. Mol. Biol.*, 41, 189–197.
- Walz, F.G. (1972) *Biopolymers*, 11, 2365–2379.
- Record, M.T., Jr., Lohman, T.M. and de Haseth, P. (1976) *J. Mol. Biol.*, 107, 145–158.
- Mascotti, D.P. and Lohman, T.M. (1990) *Proc. Natl. Acad. Sci. USA*, 87, 3142–3146.
- Wilson, W.D. and Lopp, I.G. (1979) *Biopolymers*, 18, 3025–3041.
- Wolfe, A., Shimer, G.H., Jr. and Meehan, T. (1987) *Biochemistry*, 26, 6392–6396.
- Record, M.T., Jr., Anderson, C.F. and Lohman, T.M. (1978) *Quart. Rev. Biophys.*, 11, 103–178.
- Manning, G.S. (1978) *Quart. Rev. Biophys.*, 11, 179–246.
- Record, M.T., Jr., Woodbury, C.P. and Lohman, T.M. (1976) *Biopolymers*, 15, 893–915.
- Anderson, C.F. and Record, M.T., Jr. (1980) *Biophys. Chem.*, 11, 353–360.
- Anderson, C.F. and Record, M.T., Jr. (1983) In Clementi, E. and Sarma, R.H. (eds.), *Structure and Dynamics: Nucleic Acids and Proteins*. Adenine Press, Guilderland NY, pp. 301–319.
- Olmsted, M.C., Anderson, C.F. and Record, M.T., Jr. (1991) *Biopolymers*, 31, 1593–1604.
- Bujalowski, W. and Lohman, T.M. (1987) *Biochemistry*, 26, 3099–3106.
- Dowd, J.E. and Riggs, D.S. (1965) *J. Biol. Chem.*, 240, 863–869.
- Eisenthal, R. and Cornish-Bowden, A. (1974) *Biochem. J.*, 139, 715–720.
- Cornish-Bowden, A. and Eisenthal, R. (1974) *Biochem. J.*, 139, 721–730.
- Cornish-Bowden, A. and Eisenthal, R. (1978) *Biochim. Biophys. Acta*, 523, 268–272.